Molecular Characterization of a Mouse Genomic Element Mobilized by Advanced Glycation Endproduct Modified–DNA (AGE-DNA)

Tatiana Pushkarsky,* Linda Rourke,* Lori A. Spiegel,* Michael F. Seldin,† and Richard Bucala*

*The Picower Institute for Medical Research, Manhasset, New York, U.S.A.

†Rowe Program in Genetics, Departments of Biological Chemistry and Medicine, University of California at Davis, Davis, California, U.S.A.

ABSTRACT

Background: DNA modified by advanced glycation endproducts (AGEs) undergoes a high frequency of insertional mutagenesis. In mouse lymphoid cells, these mutations are due in part to the transposition of host genomic elements that contain a DNA region homologous to the Alu family of repetitive elements. One particular 853 bp insertion, designated INS-1, was identified previously as a DNA element common to plasmids recovered from multiple, independent lymphoid cell transfections.

Materials and Methods: To characterize the genomic origin of this element, we used a 281-bp region of non-*Alu*-containing INS-1 sequence, designated *CORE*, as a probe in Southern hybridization and for screening a bacteriophage mouse genomic DNA library. The resultant clones were sequenced and localized within the mouse genome.

Results: Two distinct genomic clones of 15 kB and 17 kB in size were isolated. A 522-bp unique region common to INS-1 and corresponding to the *CORE* sequence was identified in each clone. In both cases, *CORE* was found to be surrounded by repetitive DNA sequences: a 339-bp MT repeat at the 5′ end, and a 150-bp B1 repeat at the 3′ end. The *CORE* sequence was localized to mouse chromosome 1.

Conclusions: These studies revealed that the *CORE* region of INS is present in low copy number but is associated with known repetitive DNA elements. The presence of these repetitive elements may facilitate the transposition of *CORE* by recombination or other, more complex rearrangement events, and explain in part the origin of AGE-induced insertional mutations.

INTRODUCTION

Reducing sugars such as glucose react nonenzymatically with amine-containing nucleotides to form stable products, termed advanced glycosylation endproducts (AGEs). AGEs have been shown to cause mutations in a variety of model systems (1–5). In bacteria, the mutagenic effect of AGEs results in part from the activation of host-derived, transposable elements such as IS-1 and γδ (4,5). The transposition of γδ in bacterial plasmids has been shown to occur after the modification of plasmid DNA by AGEs in vitro, as well as in native plasmids carried by bacterial strains that accumulate high intracellular levels of glucose-6-phosphate, a reactive AGE-precursor sugar (4,5).

Mutations induced by the transfection of AGE-modified DNA have also been observed in mammalian cells (1,6). In experiments in which AGE-modified plasmid DNA carrying the *lacI* mutagenesis marker was introduced into the murine myeloid cell line X63Ag8.653, the mutation rate correlated with the degree of chemical

Address correspondence and reprint requests to: Dr. Richard Bucala, The Picower Institute for Medical Research, 350 Community Drive, Manhasset, NY 11030, U.S.A. Phone: 516-562-9406; Fax: 516-365-5286; E-mail: rbucala@picower.edu

© 1997, THE PICOWER INSTITUTE PRESS. All rights reserved.
Molecular Medicine, Volume 3, Number 11, November 1997 740-749

modification by AGEs. Restriction analysis of the *lacI* gene revealed a predominance of DNA insertions which, upon subcloning and Southern hybridization, were found to be primarily of host genomic origin. DNA-sequencing analysis revealed that a large fraction (12.9%) of these insertions were due to a single, 853-bp element, termed INS-1. The structure of INS-1 was found to be that of the 180-bp, B1-repeat belonging to the *Alu* repetitive DNA family juxtaposed to a 522-bp low copy number, or unique sequence fragment. This 522-bp fragment was in turn designated *CORE*, as it was found to be juxtaposed to additional repetitive elements present within other, larger insertional mutations of the *lacI* gene that were subsequently cloned and analyzed (i.e., INS-2 and INS-3) (1).

AGE modifications accumulate in tissues as a consequence of aging and have been proposed to play a role in the development of age-related chromosomal aberrations and oncogenesis (1–7). To better understand the specific features that make the INS family of elements susceptible to mobilization by AGE-modified DNA, we cloned the mouse genomic copy of INS and characterized its genomic structure, distribution, and prevalence in different cell types. We show that the INS family of DNA elements originates from a low repetitive, genomic copy of *CORE* that lies adjacent to clusters of highly repetitive sequences. We did not observe evidence for *CORE* transposition in a variety of transformed mouse cell types, which suggests that this element may undergo mobilization only under the specific conditions associated with AGE modification.

MATERIALS AND METHODS

Cells

The mouse cell lines X63Ag8.653 (myeloma), 70Z/3 (pre-B lymphocyte), 3T3 (embryonic fibroblast), SV-T2 (SV40-transformed 3T3 fibroblasts), M-MSV (sarcoma virus–transformed 3T3 fibroblast), BNLSVA.8 (SV40-transformed liver cell), LA-4 (lung urethane–induced adenoma), and RAG (renal adenocarcinoma), were obtained from the American Type Culture Collection (Rockville, MD) and were cultured according to the supplier's recommendations.

Southern Hybridization Analyses

Genomic DNA was extracted from the liver and spleen of BALB/c and C3H mice and from mouse cell lines following a standard protocol (8). Twenty micrograms of DNA was digested to completion with the indicated restriction enzymes and electrophoresed in 0.7% agarose gels, followed by transfer to a Hybond nylon membrane (Amersham, Little Chalfont, Buckinghamshire, U.K.). After UV cross-linking, the membranes were prehybridized in "Quikhybe" solution (Stratagene, La Jolla, CA) for 30 min at 60°C. A radiolabeled DNA probe was then added and allowed to hybridize for 1–3 hr at 65°C. Membranes were washed with 1 × SSC/0.1% SDS buffer for 30 min at room temperature, followed by a 30-min wash with 0.1 × SSC/0.1% SDS buffer at 60°C. The membranes were autoradiographed with Kodak XAR film at −80°C.

Genomic Library Screening

A mouse (strain 129SV) liver genomic DNA library in the λ-FIX II vector was purchased from Stratagene. The average insert size of this library is 15 kB. A total of 5×10^6 plaques was analyzed initially by hybridization at high density (5×10^4 plaque-forming units per 150-mm plate) with a 281-bp probe corresponding to the *CORE* sequence of INS-1 (nucleotides 461–742 of the INS-1 element) (1,6) (Fig. 1). Prehybridization and hybridization were conducted at 42°C in hybridization solution (Gibco BRL, Gaithersburg, MD). The membranes were washed extensively in 0.5 × SSC/0.1% SDS at 50°C. Subsequent screenings and plaque purification were performed at lower densities under the same hybridization and washing conditions. The DNA hybridization probe was a 281-bp fragment obtained by PCR from mouse genomic DNA using primers #1 and #2 (Table 1) and corresponded to nucleotides 461–742 of the INS-1 element (1). The probe was purified using the Wizard PCR fragment purification clean kit (Promega, Madison, WI) and radiolabeled with [α-^{32}P]dCTP (3000 Ci/mmol) by using the random primer labeling kit (Amersham). Phage DNA was prepared from clones following established procedures (8), and analyzed by restriction-enzyme digestion and Southern hybridization.

DNA Sequence Analysis

Two genomic clones that were 15 kB and 17 kB in size were isolated and designated λ-1 and λ-2, respectively. The clone λ-1 was digested with *Kpn1* endonuclease, yielding a 3.7-kB DNA fragment that then was subcloned into the pUC19

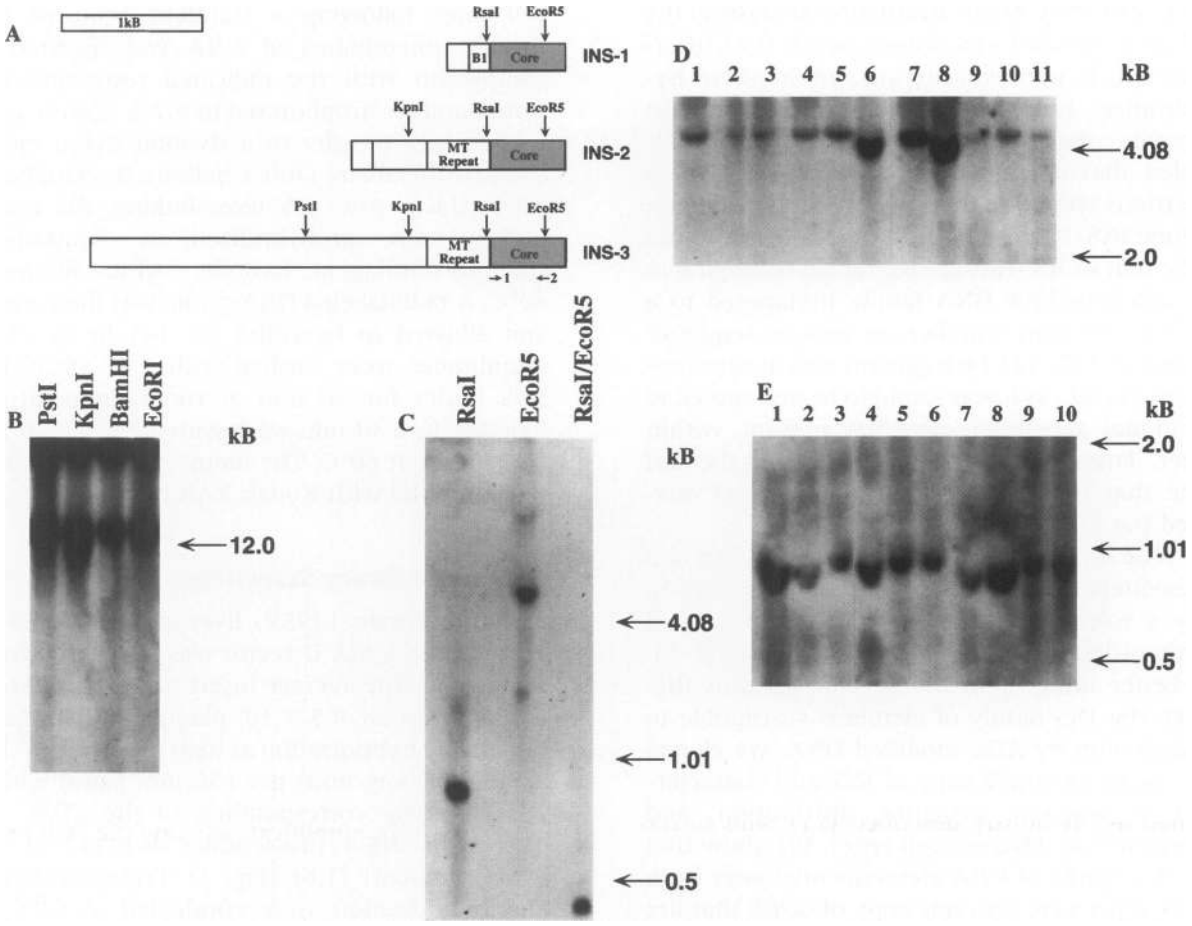


FIG. 1. Genomic organization of mouse INS elements

(A) Schematic representation of the cloned INS elements (modified from ref. 1). Primers #1 and #2 amplify a 281 nucleotide region of *CORE* which was used as the probe in panels B and C. (B, C) Mouse genomic DNA isolated from the liver of BALB/c mice (B) or from P3X63Ag8.653 lymphoid cells (C) was digested with the indicated restriction endonucleases and hybridized to a ^{32}P-labeled *CORE*-specific DNA probe. The positions of the molecular weight markers are shown on the right. (D, E) Genomic DNA isolated from spleen (lane 1) and liver (lane 2) of BALB/c mice, liver of C3H mice (lane 3), or mouse cell lines P3X63Ag8.653 (lane 4), 70Z/3 (lane 5), 3T3 (lane 6), SV-T2 (lane 7), M-MSV (lane 8), BNL SV A.8 (lane 9), LA-4 (lane 10), and RAG (lane 11) was digested with *Eco*R5 (D) or *RsaI* (E) and hybridized to a ^{32}P-labeled DNA probe shown in panel A. Minor differences in the strength of hybridization signals are due to slight variations in the amount of DNA loaded. No additional differences between lanes was observed in any part of the gel upon longer exposure. Positions of molecular weight markers are shown on the right.

vector for sequencing. The clone λ-2 was digested with *PstI* to produce a 3-kB fragment, which was subcloned into the plasmid pUC18 (9). For additional analysis, PCR amplification with primer set G/AR (Table 1) was performed, producing a single 1,456-bp fragment from each clone. These fragments were subcloned into the pT7 vector. Genomic DNA fragments were sequenced bidirectionally using the *Taq* Dye-DeoxyTerminator Cycle sequencing kit and an ABI Model 375A sequencer (Applied Biosystems, Foster City, CA). DNA sequence homology and sequence analysis were performed by using the FASTALIGN and FASTA programs, respectively (10).

PCR Analysis

The primers sets J/ALU, and G/AR (Table 1) were designed according to the sequence of INS-3 (1) and used for PCR analysis of mouse genomic DNA. The PCR amplification conditions were: 30 sec at 96°C, 30 sec at 55°C, and 30 sec at 72°C. This was performed for 35 cycles and

TABLE 1. PCR primers

Primer	Orientation	Sequence
#1	Sense	5′-AGAGTAGAGGCTCAAACCT-3′
#2	Antisense	5′-AACATCTTTGCCCAGAG-3′
G	Sense	5′-TTCTAGGACACTCTGGGT-3′
AR	Antisense	5′-TCACTCTCTAGCTGATGC-3′
J	Sense	5′-CCATCACAGATGCGTGTT-3′
ALU	Antisense	5′-TCACTCTCTAGCTGATGC-3′
I	Sense	5′-GTGTCACAGTCCTTCTAC-3′
K	Sense	5′-CGAACCTCTCTGAGTTCA-3′

was followed by a 5-min extension at 72°C. All DNA amplification reactions were performed in a GeneAmp PCR System 9600 (Perkin Elmer Cetus).

Chromosome Localization

C3H/HeJ-gld and *Mus spretus* (Spain) mice and [(C3H/HeJ-*gld* × *Mus spretus*)F1 × C3H/HeJ-*gld*] interspecific backcross mice were bred and maintained as previously described (11). *Mus spretus* was chosen as the second parent in this cross because of the relative ease of detection of informative restriction fragment length variants (RFLV) in comparison with crosses using conventional inbred laboratory strains.

DNA isolated from mouse organs by standard techniques was digested with restriction endonucleases and 10-μg samples were electrophoresed in 0.8% agarose gels. DNA was transferred to Nytran membranes (Schleicher & Schuell, Keene, NH), hybridized at 65°C with probes labeled by the random primed method with [^{32}P]-dCTP, and washed under stringent conditions, as described previously (12). The probe used for the linkage analysis included the 281-nucleotide probe corresponding to the internal sequence of *CORE*. Gene order was determined by analyzing all haplotypes and minimizing crossover frequency between all genes that were determined to be within a linkage group (12a). This method resulted in determination of the most likely gene order (13).

RESULTS

To begin to characterize the genomic structure of the sequence elements responsible for AGE-induced insertional mutagenesis, we performed Southern hybridization analysis of mouse genomic DNA using as a probe a 281-bp fragment derived from the 522-bp, low copy number region that is common to the 3′ end of the INS elements (Fig. 1A). This region, designated *CORE*, corresponds to nucleotides 329–853 of the originally described INS-1 element and borders repetitive DNA sequences at its 5′ end: the B1 repeat in INS-1, and the MT-repeat in INS-2 and INS-3.

The 281-bp hybridization probe was produced by DNA amplification using the PCR primers #1 and #2 (see Table 1). A Southern blot of a *Pst1*, *Kpn1*, *Bam*H1, and *Eco*R1 endonuclease digest of mouse genomic DNA showed a high-molecular-weight band of approximately 14 kB (Fig. 1B), indicating that relatively few copies of the *CORE* element are present in this restriction site–poor region of the mouse genome. Digestion with the *Rsa1* and *Eco*R5 endonucleases produced distinct fragments of approximately 0.9 kB and 4.5 kB, respectively, and a double digestion with *Rsa1* and *Eco*R5 gave rise to a fragment of 0.37 kB, as predicted from the known sequence of INS-1 (Fig. 1C).

We next investigated the genomic localization of *CORE* in a variety of transformed mouse cell lines. Genomic DNA was digested with both *Eco*R5 and *Rsa1* and hybridized to the 281-bp probe derived from the *CORE* sequence. Each DNA sample produced a strong hybridization signal at approximately 4.5 kB for the *Eco*R5 digest (Fig. 1D) and 0.9 kB for the *Rsa1* digest (Fig. 1E). An identical hybridization pattern was observed with the X63Ag8.653 cell line (Fig. 1C), which was used for the initial studies of AGE-induced mutagenesis (1,6). On the basis of these results, the *CORE* element appears to be a low copy number sequence within the mouse genome. These data also suggest that in all the cell lines that

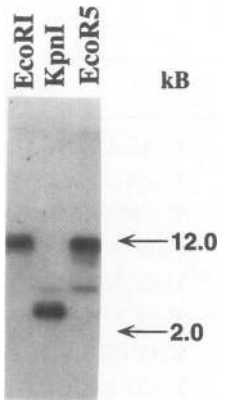


FIG. 2. Southern blotting analysis of *CORE*-containing λ-1 phage DNA

DNA isolated from the λ-1 phage containing the genomic copy of INS was digested with the indicated restriction endonucleases and hybridized to a ^{32}P-labeled *CORE*-specific DNA probe (as shown in Fig. 1A).

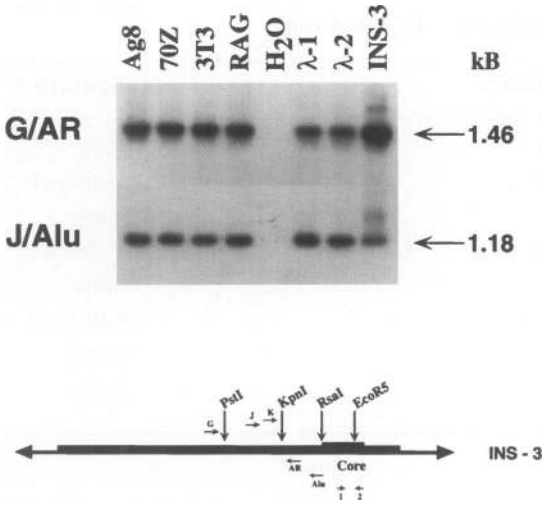


FIG. 3. PCR analysis of the genomic organization of INS elements

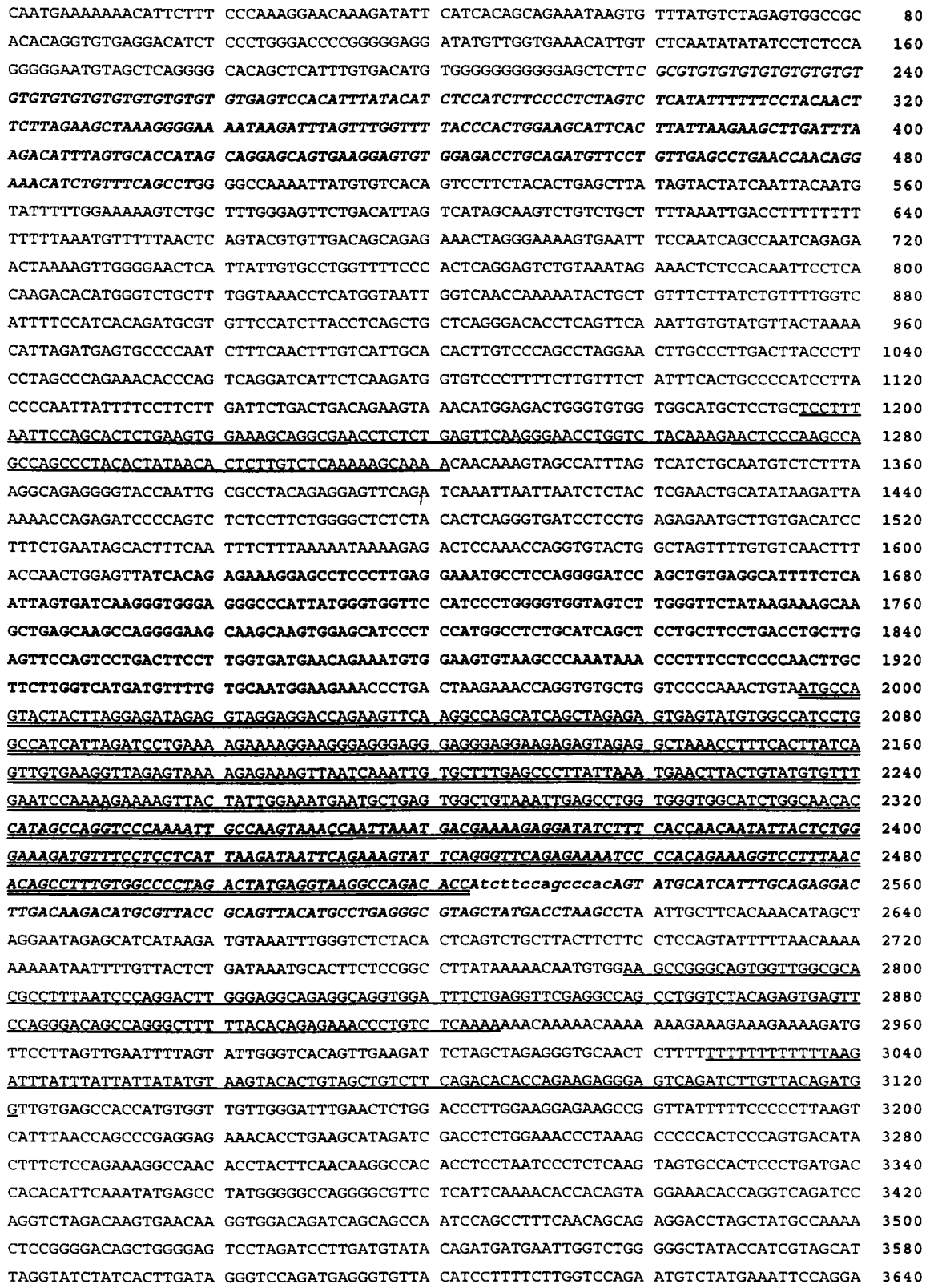
Total DNA was extracted from transformed murine cell lines (P3X63Ag8.653, 70Z/3, 3T3, RAG) or from clones containing genomic copies of INS elements (λ-1 and λ-2) and analyzed by PCR using primer pairs G/AR and J/Alu designed according to the sequence of INS-3 element. A positive PCR control was INS-3 DNA, and no DNA was added to a negative control sample (H_2O). The relative positions of the restriction enzyme sites and the PCR primer sequences are shown in the schematic for INS-3. Amplification products were revealed by Southern transfer and hybridization to a ^{32}P-labeled oligonucleotide (K).

were examined, the *CORE* sequence is localized in a region of the genome that is not readily altered by cellular transformation.

Utilizing as a hybridization probe the 281-bp *CORE* sequence, we next cloned two genomic DNA fragments from a mouse liver genomic DNA library constructed in bacteriophage λ. A 15-kB clone, designated λ-1, was found to contain the *CORE* element in a position that could be conveniently excised by the *Kpn1* endonuclease (Fig. 2). This 3.7-kB genomic *Kpn1* fragment was then subcloned into pUC18 and sequenced bidirectionally using primers #1 and #2 (Table 1). A 17-kB clone, λ-2, also contained *CORE*, but it showed a different restriction enzyme digestion pattern from the λ-1 clone: *Kpn1* digestion of the λ-2 clone produced a *CORE*-positive band that was approximately 4.5 kB (data not shown). The *CORE*-positive, 3-kB *Pst1* fragment of λ-2 was subcloned into the pUC19 plasmid and the *CORE* region was sequenced bidirectionally. This analysis revealed a 99.5% homology to the *CORE* sequence contained within the INS-1, INS-2, and INS-3 elements (1,6) (Fig. 1A and data not shown). To investigate the extent of this homology and the genomic sequence outside the *CORE* element, we next performed PCR analysis using primers located in different regions 5′ of *CORE* in INS-3 (Fig. 3) using primers G/AR and J/ALU (Table 1). DNA from several mouse cell lines and from the λ-1 genomic clone produced amplification products of the same approximate size (Fig. 3), supporting the possibility that these clones represent genomic copies of the INS-3 DNA element.

To analyze the regions of the genomic copy of INS-3 lying 5′ from the *Kpn1* site of the λ-1 clone and the *Pst1* site of the λ-2 clone—and which were omitted by the subcloning of these fragments (Fig. 1A)—we subcloned the G/AR amplification products of the λ-1 and λ-2 DNAs into pT7 vector DNA. Sequence analysis revealed a 98% identity over 2511 bp between λ-1, λ-2, and INS-3 (Fig. 4 and data not shown).

The genomic copy of the INS-3 element showed several distinct features. Three B1-like repeats (B1.1 to B1.3, Fig. 4, 5) are clustered in close proximity to the *CORE* sequence; the first is 53% homologous to the consensus B1 sequence located at position 1195–1321 of the genomic INS-3 element, the second is 68.2% homologous to B1 and is at position 2778–2926, and the third is 91.1% homologous to B1 and is located at position 3015–3121. Another feature of the



CAATGAAAAAAACATTCTTT CCCAAAGGAACAAAGATATT CATCACAGCAGAAATAAGTG TTTATGTCTAGAGTGGCCGC 80
ACACAGGTGTGAGGACATCT CCCTGGGACCCCGGGGGAGG ATATGTTGGTGAAACATTGT CTCAATATATATCCTCTCCA 160
GGGGGAATGTAGCTCAGGGG CACAGCTCATTTGTGACATG TGGGGGGGGGGGAGCTCTTC GCGTGTGTGTGTGTGTGTGT 240
GTGTGTGTGTGTGTGTGTGT GTGAGTCCACATTTATACAT CTCCATCTTCCCCTCTAGTC TCATATTTTTTCCTACAACT 320
TCTTAGAAGCTAAAGGGGAA AATAAGATTTAGTTTGGTTT TACCCACTGGAAGCATTCAC TTATTAAGAAGCTTGATTTA 400
AGACATTTAGTGCACCATAG CAGGAGCAGTGAAGGAGTGT GGAGACCTGCAGATGTTCCT GTTGAGCCTGAACCAACAGG 480
AAACATCTGTTTCAGCCTGG GGCCAAAATTATGTGTCACA GTCCTTCTACACTGAGCTTA TAGTACTATCAATTACAATG 560
TATTTTTGGAAAAAGTCTGC TTTGGGAGTTCTGACATTAG TCATAGCAAGTCTGTCTGCT TTTAAATTGACCTTTTTTTT 640
TTTTTAAATGTTTTTAACTC AGTACGTGTTGACAGCAGAG AAACTAGGGAAAAGTGAATT TCCAATCAGCCAATCAGAGA 720
ACTAAAAGTTGGGGAACTCA TTATTGTGCCTGGTTTTCCC ACTCAGGAGTCTGTAAATAG AAACTCTCCACAATTCCTCA 800
CAAGACACATGGGTCTGCTT TGGTAAACCTCATGGTAATT GGTCAACCAAAAATACTGCT GTTTCTTATCTGTTTTGGTC 880
ATTTTCCATCACAGATGCGT GTTCCATCTTACCTCAGCTG CTCAGGGACACCTCAGTTCA AATTGTGTATGTTACTAAAA 960
CATTAGATGAGTGCCCCAAT CTTTCAACTTTGTCATTGCA CACTTGTCCCAGCCTAGGAA CTTGCCCTTGACTTACCCTT 1040
CCTAGCCCAGAAACACCCAG TCAGGATCATTCTCAAGATG GTGTCCCTTTTCTTGTTTCT ATTTCACTGCCCCATCCTTA 1120
CCCCAATTATTTTCCTTCTT GATTCTGACTGACAGAAGTA AACATGGAGACTGGGTGTGG TGGCATGCTCCTGCTCCTTT 1200
AATTCCAGCACTCTGAAGTG GAAAGCAGGCGAACCTCTCT GAGTTCAAGGGAACCTGGTC TACAAAGAACTCCCAAGCCA 1280
GCCAGCCCTACACTATAACA CTCTTGTCTCAAAAAGCAAA ACAACAAAGTAGCCATTTAG TCATCTGCAATGTCTCTTTA 1360
AGGCAGAGGGGTACCAATTG CGCCTACAGAGGAGTTCAGA TCAAATTAATTAATCTCTAC TCGAACTGCATATAAGATTA 1440
AAAACCAGAGATCCCCAGTC TCTCCTTCTGGGGCTCTCTA CACTCAGGGTGATCCTCCTG AGAGAATGCTTGTGACATCC 1520
TTTCTGAATAGCACTTTCAA TTTCTTTAAAAATAAAAGAG ACTCCAAACCAGGTGTACTG GCTAGTTTTGTGTCAACTTT 1600
ACCAACTGGAGTTATCACAG AGAAAGGAGCCTCCCTTGAG GAAATGCCTCCAGGGGATCC AGCTGTGAGGCATTTTCTCA 1680
ATTAGTGATCAAGGGTGGGA GGGCCCATTATGGGTGGTTC CATCCCTGGGGTGGTAGTCT TGGGTTCTATAAGAAAGCAA 1760
GCTGAGCAAGCCAGGGGAAG CAAGCAAGTGGAGCATCCCT CCATGGCCTCTGCATCAGCT CCTGCTTCCTGACCTGCTTG 1840
AGTTCCAGTCCTGACTTCCT TGGTGATGAACAGAAATGTG GAAGTGTAAGCCCAAATAAA CCCTTTCCTCCCCAACTTGC 1920
TTCTTGGTCATGATGTTTTG TGCAATGGAAGAAACCCTGA CTAAGAAACCAGGTGTGCTG GTCCCCAAACTGTAATGCCA 2000
GTACTACTTAGGAGATAGAG GTAGGAGGACCAGAAGTTCA AGGCCAGCATCAGCTAGAGA GTGAGTATGTGGCCATCCTG 2080
GCCATCATTAGATCCTGAAA AGAAAAGGAAGGGAGGGAGG GAGGGAGGAAGAGAGTAGAG GCTAAACCTTTCACTTATCA 2160
GTTGTGAAGGTTAGAGTAAA AGAGAAAGTTAATCAAATTG TGCTTTGAGCCCTTATTAAA TGAACTTACTGTATGTGTTT 2240
GAATCCAAAAGAAAAGTTAC TATTGGAAATGAATGCTGAG TGGCTGTAAATTGAGCCTGG TGGGTGGCATCTGGCAACAC 2320
CATAGCCAGGTCCCAAAATT GCCAAGTAAACCAATTAAAT GACGAAAAGAGGATATCTTT CACCAACAATATTACTCTGG 2400
GAAAGATGTTTCCTCCTCAT TAAGATAATTCAGAAAGTAT TCAGGGTTCAGAGAAAATCC CCACAGAAAGGTCCTTTAAC 2480
ACAGCCTTTGTGGCCCCTAG ACTATGAGGTAAGGCCAGAC ACCatcttccagcccacAGT ATGCATCATTTGCAGAGGAC 2560
TTGACAAGACATGCGTTACC GCAGTTACATGCCTGAGGGC GTAGCTATGACCTAAGCCTA ATTGCTTCACAAACATAGCT 2640
AGGAATAGAGCATCATAAGA TGTAAATTTGGGTCTCTACA CTCAGTCTGCTTACTTCTTC CTCCAGTATTTTTAACAAAA 2720
AAAAATAATTTTGTTACTCT GATAAATGCACTTCTCCGGC CTTATAAAAACAATGTGGAA GCCGGGCAGTGGTTGGCGCA 2800
CGCCTTTAATCCCAGGACTT GGGAGGCAGAGGCAGGTGGA TTTCTGAGGTTCGAGGCCAG CCTGGTCTACAGAGTGAGTT 2880
CCAGGGACAGCCAGGGCTTT TTACACAGAGAAACCCTGTC TCAAAAAAACAAAAACAAAA AAAGAAAGAAAGAAAAGATG 2960
TTCCTTAGTTGAATTTTAGT ATTGGGTCACAGTTGAAGAT TCTAGCTAGAGGGTGCAACT CTTTTTTTTTTTTTTAAG 3040
ATTTATTTATTATTATATGT AAGTACACTGTAGCTGTCTT CAGACACACCAGAAGAGGGA GTCAGATCTTGTTACAGATG 3120
GTTGTGAGCCACCATGTGGT TGTTGGGATTTGAACTCTGG ACCCTTGGAAGGAGAAGCCG GTTATTTTTCCCCCTTAAGT 3200
CATTTAACCAGCCCGAGGAG AAACACCTGAAGCATAGATC GACCTCTGGAAACCCTAAAG CCCCCACTCCCAGTGACATA 3280
CTTTCTCCAGAAAGGCCAAC ACCTACTTCAACAAGGCCAC ACCTCCTAATCCCTCTCAAG TAGTGCCACTCCCTGATGAC 3340
CACACATTCAAATATGAGCC TATGGGGGCCAGGGGCGTTC TCATTCAAAACACCACAGTA GGAAACACCAGGTCAGATCC 3400
AGGTCTAGACAAGTGAACAA GGTGGACAGATCAGCAGCCA ATCCAGCCTTTCAACAGCAG AGGACCTAGCTATGCCAAAA 3500
CTCCGGGGACAGCTGGGGAG TCCTAGATCCTTGATGTATA CAGATGATGAATTGGTCTGG GGGCTATACCATCGTAGCAT 3580
TAGGTATCTATCACTTGATA GGGTCCAGATGAGGGTGTTA CATCCTTTTCTTGGTCCAGA ATGTCTATGAAATTCCAGGA 3640

FIG. 4. Sequence of the genomic copy of the murine INS element derived from the λ-1 genomic DNA clone

The regions of INS homologous to B-1 repeats are underlined, the region homologous to microsatellite DNA is italicized, the region homologous to human Sau3A repeat is in bold and italicized, and the region homologous to the MT-repeat of c-*abl* oncogene is in bold. *CORE* is underlined by the double line, and the 13-bp element homologous to mouse minisatellite DNA is indicated by lower case letters. The region homologous to INS-2 is located between nucleotides 1291 and 2521. Primers #1 and #2 are at positions 2132–2151 and 2394–2411, respectively.

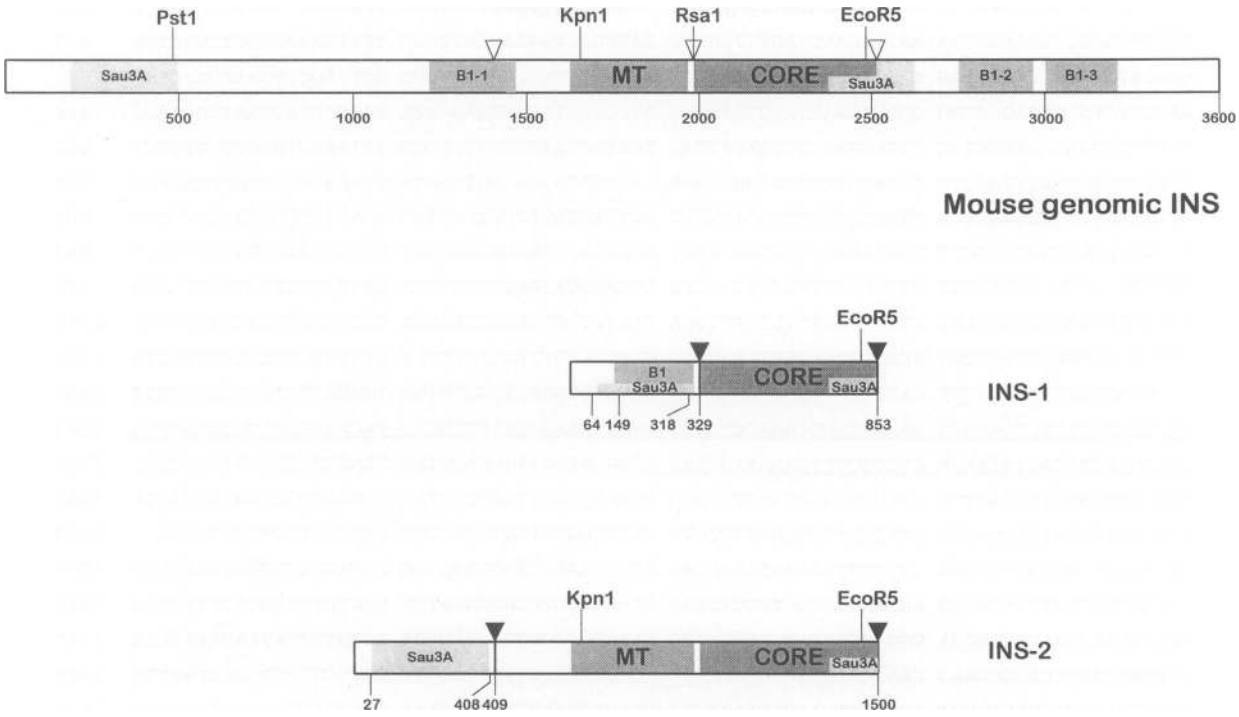


FIG. 5. Schematic representation of colocalization of repetitive DNA with the recombination breakpoints in INS-1 and INS-2, and the genomic copy of INS-3

The class of repetitive DNA is designated in boxes, filled arrowheads show the actual recombination breakpoints in mutant DNA (INS-1 and INS-2), open arrowheads show the presumed breakpoints in the genomic copy of INS-3.

INS-3 element is a 86.6% homology of the 338-bp fragment located at the 5′ end (position 1615–1953) to the MT repeat (13a) of the mouse *c-abl* oncogene (Figs. 4, 5). Of note is the finding that both the 5′ and the 3′ portions of genomic INS-3 (nucleotides 240–497 and 2332–2620) show a 44.2% and a 43.5% homology, respectively, to the human Sau3A repeat (14,15). Both stretches of Sau3A homology are in the sense orientation. Sequence analysis of INS-1 and INS-2 also revealed a 254-bp and 384-bp tract of DNA, respectively, that showed 40% homology with the Sau3A repeat. Both of these Sau3-like regions were located 5′ to the *CORE* "breakpoint." In INS-1, the Sau3A-homologous region extends for 250 bp between nucleotides 64 and 318 in an inverted direction and precedes the 5′ breakpoint (nucleotide 329) by 11 nucleotides. In INS-2, the Sau3A-homologous region is located between nucleotides 27 and 408, is also in an inverted position, and is located adjacent to the breakpoint at position 409 (Fig. 5).

Finally, to determine the chromosomal location of the mouse INS-derived *CORE* sequence, we analyzed a panel of DNA samples from an interspecific backcross that has been characterized for more than 1200 genetic markers throughout the genome. The genetic markers included in this map span between 50 and 80 centi-Morgans on each of the mouse autosomes and the X chromosome. Initially, DNA from the two parental mice [C3H/HeJ-*gld* and (C3H/HeJ-*gld* × *Mus spretus*)F1] was digested with various restriction endonucleases and hybridized with the 281-nucleotide probe corresponding to the *CORE* element to determine restriction fragment length variants (RFLVs) for haplotype analyses. Informative *Taq*1 RFLVs were detected: C3H/HeJ-*gld*, 5.2 kB; *Mus spretus*, 5.4 kB.

Comparison of the haplotype distribution of the *CORE* RFLVs indicated that the *CORE* locus cosegregated in all of the 114 meiotic events examined with the renin gene locus (Ren1 and Ren2) on mouse Chr1 (Fig. 6). The haplotype distribution among other genes localized to mouse Chr1 is shown in Figure 6. The best gene order (13) ± the standard deviation (12a) was determined to be (centromere) C4bp-1.8 cM ± 1.2 cM-*CORE*/Ren1/Ren2-7.0 cM ± 2.4 cM-

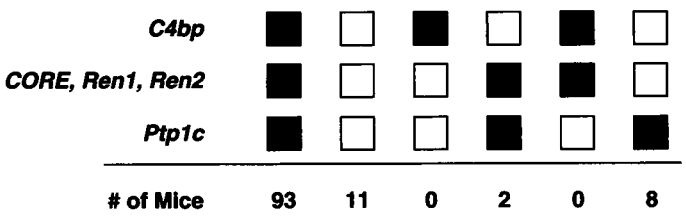


FIG. 6. Segregation of *CORE* on mouse chromosome 1 in [(C3H/HeJ-*gld* × *Mus spretus*)F1 × C3h/HeJ-*gld*] interspecific backcross mice

Filled boxes represent the homozygous C3H pattern and open boxes represent the F1 pattern. The abundance of C3H/HeJ haplotypes reflects the selection of this panel for gld/gld phenotype positive mice that is located on mouse chromosome 1.

Ptp1c. This linkage group corresponds to the 1q32 band in the human chromosome.

DISCUSSION

We previously showed that sequences contained within the mouse genome are mobilized after transfection of AGE-modified DNA, resulting in the insertion of host elements into target plasmids (1,6). Detailed analysis of the DNA insertions that were obtained from multiple, independent transfections led to the cloning of three elements INS-1, INS-2, and INS-3 which differed in their overall lengths but which were 99.9% homologous over 522 bp of their respective 3′ ends. This 522-bp homologous fragment was designated *CORE*. The 5′ end of *CORE* is adjacent to the B1 repeat in INS-1 (1,6) and to the middle repetitive MT element (13a) in INS-2 and INS-3. To gain insight into both the origin and the organization of INS element(s) in the mouse genome and the potential mechanism of their transposition, we performed Southern hybridization analysis of mouse DNA isolated from various transformed cell lines using as a probe a 281-bp fragment corresponding to *CORE*. This analysis revealed that INS is a low copy number element associated with only a minimal degree of restriction fragment length polymorphism. Two recombinant clones carrying genomic INS copies were isolated, partially sequenced, and found to have 97.5% homology with each other and with the INS-3 element (covering 2511 bp, including the *CORE* sequence). In all, we determined 3640 bp of the genomic sequence corresponding to the host-derived, INS element. This fragment was found to contain a cluster of three B1-like sequences: the 3′-most B1.3 is 91.1% homologous to the consensus B1-element, while the remaining two are more divergent: B1.1 shows 53% homology, and B1.2 shows 68.2% homology to the consensus B1 sequence (Fig. 5).

Clusters of B1-like repetitive elements have been described previously in the case of *Alu* repeats in human DNA (16), and these elements also have been localized to the 3′ and 5′ sides of chromosomal breakpoints (17,18). *Alu* repeat clusters have been implicated in the formation of the Philadelphia chromosome (Ph1), which results from the translocation of the long arms of chromosomes 9 and 22 (16–18). *Alu* sequences may facilitate recombination by forming hairpin structures, thus providing a symmetrical topology recognized by the enzymatic complexes that mediate DNA recombination (16). The insertion of INS sequences into AGE-modified plasmid DNA may occur by similar recombination events mediated by B1 repeats. Since these repeats are in close proximity to *CORE*, recombination could lead to a series of *CORE*-containing elements with differing 3′ and 5′ ends, thus accounting for the structures of the DNA insertions observed in these studies.

DNA sequencing analysis revealed that the 3′ breakpoint of each of the three plasmid-derived INS insertions (INS-1, INS-2, and INS-3) was fixed, and occurred at nucleotide 2518 of the genomic INS sequence. INS-1 and INS-2 appear to be formed by recombination of the B1 repeat "B1.3" with the repetitive DNA upstream of the *CORE* (either the MT repeat, or B1). The peculiar structure of the 5′ ends of INS-1 and INS-2, which contain repetitive sequences that were not found in the genomic copy of INS, may then result from additional, downstream recombination events.

Another interesting feature of the genomic INS element is the 13-nucleotide sequence located at the 3′ breakpoint (nucleotides 2514 to 2527). This sequence displays 91.7% homology to the repetitive unit of mouse minisatellite DNA. Minisatellite DNA consists of tandemly repeated sequences that are genetically unstable and often gain or lose their repeat units. The main repeat unit, GGAGGTGGGCAGGA(G/A)G,

has been shown to serve as a binding site for proteins involved in recombination, which suggests that this region also may act as a binding site for a recombination enzyme(s) in the case of INS elements (19).

The position of *CORE* on mouse chromosome 1 strongly suggests that the human orthologue of this gene is positioned on human chromosome 1 band q31 (20,21,21a). Using probes derived from mouse INS sequences, we failed to find a strong hybridization signal for *CORE* or other INS regions in Southern hybridization analysis of human genomic DNA. Low-stringency hybridization using the 281-bp *CORE* fragment as a probe did, however, produce an interspersed pattern of hybridization signals. It is noteworthy that a Genebank search of the human genome for sequences related to mouse INS revealed two regions of homology to the human alphoid Sau3A family of repetitive DNA: a 257-bp fragment located at nucleotides 241–447, showing 43.5% homology, and a 290-bp segment at nucleotides 2320–2610, with 44.2% homology (Fig. 5). The Sau3A repetitive DNA family consists of tandemly repeated, 170-bp units that exhibit 68–96% homology between individual elements. Alphoid Sau3A repeats have been mapped to the centromeric region of several human chromosomes and have been implicated in the formation of extrachromosomal self-replicating DNA structures (14,15). It is conceivable that mouse INS sequences may be involved in these functions in a manner similar to those ascribed for human Sau3A repeats (14,15).

The products of advanced glycation chemically modify nucleotide bases, producing covalent adducts such as carboxyethylguanine (22). Stable crosslinks involving both protein and DNA also may result from glycation events and it may be difficult for cells to repair these lesions by means other than by DNA transposition or recombination (1,3,23). In all likelihood, these attempts at chromosomal repair would involve repetitive DNA sequence elements, such as those described here. Glycation damage accumulates in tissues over time and may play an important role in the DNA rearrangements that contribute to age-associated chromosomal aberrations leading to oncogenesis (1–7).

ACKNOWLEDGMENTS

This work was supported by NIH grant 2RO1 DK19655 and by NIH grant HG00734.

REFERENCES

1. Bucala R, Lee A, Rourke L, Cerami A. (1993) Transposition of an Alu-containing element induced by DNA-advanced glycosylation endproducts. *Proc. Natl. Acad. Sci. U.S.A.* **90:** 2666–2670.
2. Bucala R, Model P, Cerami A. (1984) Modification of DNA by reducing sugars: A possible mechanism for nucleic acid aging and age-related dysfunction in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **81:** 105–109.
3. Bucala R, Model P, Russel M, Cerami A. (1985) Modification of DNA by glucose-6-phosphate induces DNA rearrangements in a *Escherichia coli* plasmid. *Proc. Natl. Acad. Sci. U.S.A.* **82:** 8439–8442.
4. Lee AT, Cerami A. (1987) Elevated glucose-6-phosphate levels are associated with plasmid mutations in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **84:** 8311–8314.
5. Lee AT, Cerami A. (1991) Induction of gamma delta transposition in response to elevated glucose-6-phosphate levels. *Mutat. Res.* **249:** 125–133.
6. Pushkarskaya T, Rourke L, Spiegel L, Cerami A, Bucala R. (1994) Identification of the genomic origin of INS-1, a eukaryotic transposable element induced by DNA—Advanced glycosylation endproducts, abstr. 780. In: *Abstracts of the 85th Annual Meeting of the American Society for Biochemistry and Molecular Biology, 1994.* American Society for Biochemistry and Molecular Biology, Washington, D.C. *FASEB J.* **8:** A1393.
7. Bucala R, Cerami A. (1992) Advanced glycosylation: Chemistry, biology and implications for diabetes and aging. *Adv. Pharmacol.* **23:** 1–34.
8. Sambrook J, Fritsch EF, Maniatis T. (1990) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
9. Yanisch-Perron C, Vieira J, Messing J. (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequence of M13mp18 and pUC19 vectors. *Gene* **33:** 103.
10. Pearson WR, Lipman DJ. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* **85:** 2444–2448.
11. Seldin MF, Morse HS, Reeves JP, Scribner JP, LeBoeuf RC, Steinberg D. (1988) Genetic analysis of autoimmune *gld* mice. I. Identification of a restriction fragment length polymorphism closely linked to the gld mutation

within a conserved linkage group. *J. Exp. Med.* **167:** 688–693.

12. Watson ML, Seldin MF. (1994) Application of mouse crosses towards defining the genetics of disease phenotypes. *Methods Mol. Genet.* **5:** 369–387.

12a. Green EL. (1981) Linkage, recombination and mapping. *In* Genetics and Probability in Animal Breeding Experiments. E. Green, editor. Macmillian, New York, p. 77–113.

13. Bishop DT. (1985) The information content of phase-known mating for ordering genetic loci. *Genet. Epidemiol.* **2:** 349–361.

13a. Adiletta DC, Elliott RW, Woodworth ME. (1993) Characterization of murine middle repetitive DNA. DNA & Cell Biol. **12:** 319–327.

14. Kiyama R, Matsui H, Oishi M. (1986) A repetitive DNA family. *Proc. Natl. Acad. Sci. U.S.A.* **83:** 4665–4669.

15. Ohki R, Oishi M, Kiyama R. (1995) Preference of the recombination sites involved in the formation of extrachromosomal copies of the human alphoid Sau3A repeat family. *Nucl. Acids Res.* **23:** 4971–4977.

16. Calabretta B, Robberson DL, Barrera-Saldana HA, Lambrou TP, Saunders GF. (1982) Genome instability in a region of human DNA enriched in *Alu* repeat sequences. *Nature* **296:** 219–225.

17. Chen SJ, Chen Z, Font MP, d'Auriol L, Larsen CJ, Berger R. (1989) Structural alterations of the BAR and ABL genes in Ph1 positive acute leukemias with rearrangements in the BCR gene first intron: Further evidence implicating Alu sequences in the chromosome translocation. *Nucl. Acids Res.* **17:** 7631–7642.

18. De Klein A, van Agthoven T, Groffen C, Heisterkamp N, Groffen J, Grosveld G. (1986) Molecular analyses of both translocation products of a Philadelphia-positive CML patients. *Nucl. Acids Res.* **14:** 7071–7082.

19. Shinder GA, Manam S, Ledwith BJ, Nichols WW. (1994) Minisatellite DNA-binding proteins in mouse brain, liver, and kidney. *Exp. Cell Res.* **213:** 107–112.

20. http://saturn.nlm.nih.gov/Omim/Homology/

21. Watson ML, D'Eustachio P, Mock BA, Steinberg AD, Morse III HC, OaKey RJ, Howard TA, Rochelle JM, Seldin MF. (1992) A linkage map of mouse chromosome 1 using an interspecific cross segregating for the *gld* autoimmunity mutation. *Mammal. Genome* **2:** 158–171.

21a. DeBry RW, Seldin MF. (1996) Human/mouse homology relationships. Genomics **33:** 337–351.

22. Papoulis A, Al-Abed Y, Bucala R. (1995) Identification of N2-(1-carboxyethyl)guanine (CEG) as a guanine advanced glycosylation end product. *Biochemistry* **34:** 648–655.

23. Sargent RG, Brenneman MA, Wilson JH. (1997) Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol. Cell Biol.* **17:** 267–277.

Communicated by D. Weatherall. Accepted September 19, 1997.